Computers, Materials & Continua

Tech Science Press

DOI: 10.32604/cmc.2024.056742

ARTICLE

Constructive Robust Steganography Algorithm Based on Style Transfer

Xiong Zhang[1,2], **Minqing Zhang**[1,2,3,*], **Xu'an Wang**[1,2,3,*], **Siyuan Huang**[1,2] **and Fuqiang Di**[1,2]

[1]College of Cryptography Engineering, Engineering University of People's Armed Police, Xi'an, 710086, China

[2]Key Laboratory of People's Armed Police for Cryptology and Information Security, Engineering University of People's Armed Police, Xi'an, 710086, China

[3]Key Laboratory of CTC & Information Engineering, Ministry of Education, Engineering University of People's Armed Police, Xi'an, 710086, China

*Corresponding Authors: Minqing Zhang. Email: api_zmq@126.com; Xu'an Wang. Email: wangxazjd@163.com

Received: 29 July 2024 Accepted: 18 September 2024 Published: 15 October 2024

ABSTRACT

Traditional information hiding techniques achieve information hiding by modifying carrier data, which can easily leave detectable traces that may be detected by steganalysis tools. Especially in image transmission, both geometric and non-geometric attacks can cause subtle changes in the pixels of the image during transmission. To overcome these challenges, we propose a constructive robust image steganography technique based on style transformation. Unlike traditional steganography, our algorithm does not involve any direct modifications to the carrier data. In this study, we constructed a mapping dictionary by setting the correspondence between binary codes and image categories and then used the mapping dictionary to map secret information to secret images. Through image semantic segmentation and style transfer techniques, we combined the style of secret images with the content of public images to generate stego images. This type of stego image can resist interference during public channel transmission, ensuring the secure transmission of information. At the receiving end, we input the stego image into a trained secret image reconstruction network, which can effectively reconstruct the original secret image and further recover the secret information through a mapping dictionary to ensure the security, accuracy, and efficient decoding of the information. The experimental results show that this constructive information hiding method based on style transfer improves the security of information hiding, enhances the robustness of the algorithm to various attacks, and ensures information security.

KEYWORDS

Information hiding; neural style transfer; robustness; map dictionary

1 Introduction

Information hiding is the process of hiding secret information in a host signal in an invisible manner, and extracting the secret information when needed, in order to achieve purposes such as covert communication and copyright protection [1]. These carriers often encompass a diverse range of multimedia types, including text, images, audio, and video. Digital images, in particular, have become one of the hot topics in information hiding research due to their widespread application in network environments.

Traditional image steganography typically employs various techniques to adjust images, often modifying carrier data such as images [2], audio, and video [3,4] to implant secret information. For instance, techniques like the least significant bit (LSB) in the spatial domain [5], JPEG (Joint Photographic Experts Group) compression domain techniques [6], and transform domain methods such as Discrete Wavelet Transformation (DWT), Discrete Cosine Transform (DCT) [7,8], and Discrete Fourier Transform (DFT) [9] are used for information embedding. However, such modifications can leave detectable traces on the carrier, and once detected by professional steganographic analysis tools, this could lead to information leakage, increasing the risk of secret information being discovered. Against this backdrop, the challenge of effectively hiding information without leaving traces has become an urgent and formidable issue.

In recent years, researchers have proposed "constructive image steganography". This new type of information hiding method cleverly utilizes the properties or features of the carrier itself for information hiding without significantly changing the original carrier. It does not require any modifications to the original carrier, effectively avoiding the problem of leaving detectable traces. This method better conceals secret information and enhances the ability to resist steganalysis. The existing constructive steganography algorithms mainly utilize the mapping relationship between specific stego images and secret information to achieve covert transmission of information. This greatly improves the security of information hiding.

1.1 Related Works

Zhou et al. [10] have pioneered a carrier-free steganographic algorithm based on the mapping relationship between images and ciphertext, marking a significant advancement in carrier-free information hiding technology. Building on this foundation, further extensions have been applied to various scenarios and methods. As described in Zhou et al.'s [11] work, by employing the BOW (Bag-of-Words) model to extract visual keywords from images to represent secret information, this technique avoids direct modification of the carrier image, thus reducing the risk of detection. However, the requirement for a vast image library to construct the codebook makes it difficult to scale and limits the capacity and robustness of the hidden information. In Yuan et al.'s [12] work, an attempt was made to enhance the robustness of carrier-free information hiding by utilizing SIFT (Scale-Invariant Feature Transform) features and the BOF (Bag-of-Features) model. This strategy maps the SIFT features of an image to secret information, improving security, but still facing the issue of limited hiding capacity. Further exploration in Zhou et al.'s [13] work involved a target recognition-based hiding method, using Faster-RCNN (Faster Region-based Convolutional Neural Networks) to detect object categories in images and associate secret information with these categories. This approach improved in terms of hiding capacity and robustness but was constrained by the limited number of categories in natural images, restricting the hiding of large-scale information. Meng et al. [14] leveraged target recognition technology by analyzing the types and positions of multiple targets in an image, constructing an efficient steganographic strategy based on target detection and relational mapping. Cao et al. [15] proposed a dynamic content selection framework that achieves efficient information hiding in dynamic environments by intelligently selecting the most suitable images to represent secret information. Wang et al. [16] utilized advanced Generative Adversarial Network (GAN) technology to generate images capable of carrying high-capacity secret information, greatly enhancing the security and practicality of information hiding. Shi et al. [17] proposed disguising steganographic tools as deep neural network networks to perform style transfer tasks. In their approach, the neural network is manipulated to pass a given style to transform the image into the target style, while embedding secret data into the given image.

However, this method that relies on specific mapping relationships still has some weaknesses. Especially when stego images encounter geometric or non-geometric attacks in the cloud or during network transmission, it may cause changes in their pixel values, making it difficult for the receiver to correctly extract hidden information. To overcome these challenges, we propose an innovative style transfer based constructive image steganography method. This method can not only effectively deal with attacks during transmission, but also improve the robustness and security of information hiding. The advantage of this method is that it cannot only perform style transfer [18] while maintaining image content but also hide information in style features, providing higher security and improved robustness.

1.2 Objective of This Study

Our research objectives can be summarized as follows:

- In order to better ensure the security of information during transmission, we have designed a mapping dictionary that maps secret information to image categories. Then, through style transfer techniques, we combine the style of the secret image with the content of the content image to generate a stego image.

- In order to enhance the robustness of the algorithm, the reconstruction network we designed integrates residual blocks and dense blocks to explore the local features of the stego image more deeply and promote the memory effect on these features.

- In order to enhance the integrity of information and the anti-attack capability of secret images, we designed a training set consisting of stego images processed by various attack methods and conducted a series of attacks on these stego images before inputting them into the reconstruction network to ensure accurate recovery of secret images.

Our research aims to enhance the reliability, security, anti-attack capability, and application scope of information hiding technology to address various challenges in practical applications and ensure information security.

2 The Proposed Scheme

The core of this approach encompasses three stages: the mapping of secret messages, the generation of stego images, and the reconstruction of secret information.

In the process of mapping secret messages, we first select some images from various categories to establish a style image library and a content image library. Build a binary code mapping dictionary based on the categories of natural images in the style image library. We encode the secret information into binary digits and then map it to the secret image through a mapping table.

In the stego image generation stage, we input secret images and public images and perform joint preprocessing on them. By using neural style conversion technology, the style features of secret images are combined with the content features of content images to obtain a stego image that can naturally and effectively hide information. We can accurately reconstruct the original secret information from stego images through a secret image reconstruction network.

As shown in Fig. 1, this article describes the overall framework of the algorithm.

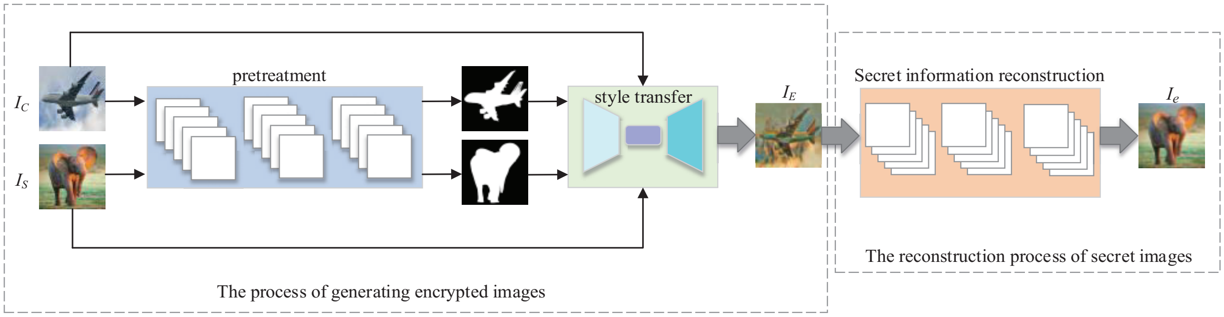

I_C
I_S
pretreatment
style transfer
I_E
Secret information reconstruction
I_e
The reconstruction process of secret images
The process of generating encrypted images

Figure 1: The overall framework of constructive robust steganography algorithm based on style transfer

In this context, the acronym I_C stands for the Content Image, I_S denotes the Style Image, which is the covert image derived from a library of categorized images and mapped to carry the binary-encoded secret message. I_E represents the style-modified stego image, where the style features of the secret image are artfully merged with the content of the public image, creating a visually coherent image that conceals the information within. Finally, I_e refers to the reconstructed secret image.

2.1 Constructing a Secret Mapping Image Library

In the information hiding algorithm proposed in this article, we need to select the corresponding image from the style image library as the secret image based on the secret information. To achieve this process, one of the essential steps is to establish a mapping relationship, that is, a corresponding dictionary between the image categories in the style image library and the binary codes. Assuming the library contains n different natural image categories, denoted as $C = \{C_1, C_2, \ldots, C_n\}$, where C_n represents the nth category image. Correspondingly, the binary code length for each category C_n is $K = \log 2^{n}$.

To illustrate this mapping relationship, we provide an example dictionary as shown in Table 1. This mapping dictionary is applied to the process of hiding and extracting secret information. Therefore, before engaging in covert communication, the sender and receiver must share this dictionary to ensure the correct transmission and interpretation of the information.

Table 1: Construct a mapping dictionary

Secret information (binary code)/K bits	Category (C_n)
00 . . . 00	Fish
00 . . . 01	Dog
00 . . . 10	Elephant
. . .	. . .
11 . . . 10	Airplane
11 . . . 11	People

2.2 The Process of Generating Stego Images

2.2.1 Preprocessing Stage

When applying the style attributes of secret images to content images, it is necessary to maintain the original content characteristics of the content images and ensure that their structure is not affected. If style transfer is directly performed on secret images and content images, it may result in image distortion or exposure of secret information. To this end, we introduced a preprocessing step that includes semantic segmentation of two types of images, followed by style transfer of the segmented images. This ensures that styles are correctly transferred according to the image regions and prevents distortion of the image structure.

The segmentation technique used is based on the method proposed in Reference [19]. We use a model similar to that proposed by Chen et al. [20] to generate image segmentation masks. After segmentation, the secret image and content image are combined with their corresponding masks and input into a style transfer network for processing. We enhance feature extraction through an adaptive convolution strategy, which can more effectively preserve detailed sections of the image.

2.2.2 Stego Image Generation Stage

In this study, we employed a robust method based on deep learning to integrate the style of secret images and the content of content images, utilizing the VGG (Visual Geometry Group)-19 [21] network architecture. We use semantically segmented images, secret images, and content images as inputs, but the difference is that our method extracts features at different stages of the network.

To focus on generating stego images, we only utilize the first five convolutional layer blocks of the VGG-19 network. These convolutional layer blocks are capable of capturing multi-scale features from shallow to deep layers. In addition, we avoided using the fully connected layers of the VGG-19 network as these layers are primarily used for classification tasks rather than feature extraction. Through this approach, we ensure that the network focuses on extracting the most relevant features.

In Fig. 2, we inputted the secret image I_S and content image I_C. In this network, the image size is uniformly adjusted to 128 × 128 pixels. In the initialization phase, we use the secret image I_S as the starting point for generating the stego image I_E. The purpose of doing so is to preserve the key visual elements of the secret image in subsequent processing.

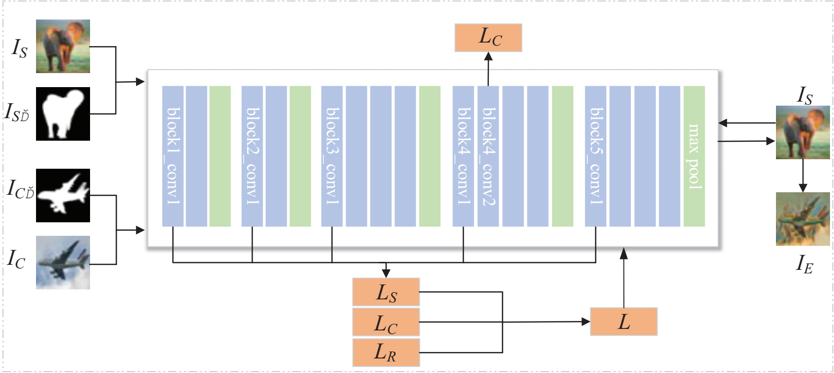


Figure 2: The process of generating the stego image

In this context, I_S denotes the Style Image, which is the secret image mapped from the image category library. $I_{S'}$ represents the segmented style image, while $I_{C'}$ signifies the segmented content image. I_C stands for the Content Image, I_E indicates the style-modified stego image, L_S is the style loss, L_C is the content loss, L_R is the reconstruction loss, and L is the total loss.

We utilize the second convolutional layer within the fourth convolutional block, conv4_2, to capture the content features of the content image I_C, as shown in Eq. (1). This layer's extracted content features are capable of representing the primary visual information of the content image I_C.

$$F_C = \frac{1}{2}\sum_{i,j}\left(F\left[I_E\right] - F\left[I_C\right]\right)^2_{i,j} \tag{1}$$

In the equation, F_C represents the content features of the content image, I_E denotes the generated stego image, I_C signifies the content image, and $F\left[\cdot\right]$ indicates the feature matrix located at position (i, j). We use the first convolutional layer of each of the first five blocks in the VGG-19 network to capture and encode the style attributes of the stego image, subtly integrating the secret information into the texture and color of the image. This allows for the transmission of hidden visual elements without drawing attention. Additionally, we employ the generated image segmentation mask as an additional channel for both the secret and content images to reinforce the preservation of style features. This method, by enhancing the conveyance of style information, helps to mitigate potential style distortion that may occur during the style transfer process. We denote the style loss by L_S, as shown in Eq. (2).

$$L_S = \sum_{C=1}^{C}\frac{1}{2N^2}\sum_{i,j}\left(G\left[I_E\right] - G\left[I_s\right]\right)^2_{i,j} \tag{2}$$

In Eq. (2), C represents the number of channels in the semantic segmentation mask, and $G[\cdot]$ is the Gram matrix of $F\,F[\cdot]$, as shown in Eq. (3). This method allows for a deeper understanding of the feature interactions within the image, enhancing the precision of our semantic segmentation and style transfer processes.

$$G\left[\cdot\right] = F\left[\cdot\right]\cdot F\left[\cdot\right]^T \tag{3}$$

Here, we employ a downsampling technique to process the segmentation mask, ensuring that its resolution matches the spatial dimensions of the feature maps at various layers of the VGG19 network. Additionally, to ensure the accuracy and clarity of the content in the final composite stego image, we have specifically designed a regularization term R. This regularization term is implemented by performing a matrix multiplication operation between the vectorized transpose matrix of the stego image and the matrix of the content image. The corresponding mathematical expression is shown in Eq. (4). In this way, the algorithm ensures a high degree of content consistency between the stego image and the content image, thus avoiding distortion.

$$H = \sum_{C=1}^{3} V_C\left[I_E\right]^T M V_C\left[I_E\right] \tag{4}$$

In the formula, C represents the number of channels, and $V_C\left[I_E\right]$ refers to the output of the stego image I_E on channel C, which is vectorized into the $N \times 1$ column vector. And M represents an $N \times N$ matrix form of the input content image I_C, where N is the total number of pixels in the image.

Specifically, the total discrepancy can be calculated by comparing the pixel value differences between the stego image and both the secret and content images across all channels. The specific formula is shown in Eq. (5).

$$L = \alpha F_C + \beta L_S + \gamma H \tag{5}$$

In this paper, the parameters α and β are defined to adjust the content and style feature weights of the stego image I_E, respectively, while γ is the weight coefficient for the regularization term. The setting of these three parameters aims to ensure that the stego image I_E is visually indistinguishable from regular images. Specifically, we select $\alpha = 1$, $\beta = 10^3$, and $\gamma = 10^4$ as the parameter values. After adjusting these parameters, the stego image I_E is updated through Eq. (5) and re-entered into the VGG network. This process is iteratively repeated until the generated stego image is consistent with the content image in terms of content and matches the secret image in style, achieving the predetermined goals of information hiding and recovery.

2.3 Secret Information Reconstruction Phase

The recipient acquires the disguised image and the set of parameters for the convolutional neural network designed for reconstruction through public communication channels. The training of this reconstruction network is based on two categories of stego images: one category consists of images that have been perturbed by potential attacks, and the other category comprises undisturbed images. Utilizing this data, the network learns how to recover the hidden information from possible attacks. When the recipient inputs the received stego image into this well-trained reconstruction network, they are able to decode and obtain the original transmitted secret image.

2.3.1 Structure of the Reconstruction Network

Considering that stego images transmitted over public communication channels may be subject to attacks such as quantization, the network structure designed in this paper must take into account the ability to successfully extract the hidden image information even after the stego images have been subjected to these common attacks.

After attempting and comparing various network architecture designs, we selected the architecture shown in Fig. 3, which demonstrated the best performance in experiments.

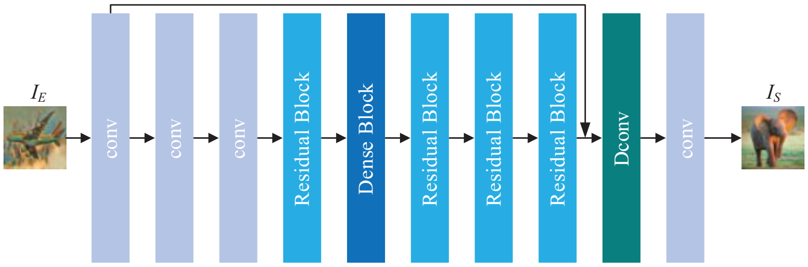


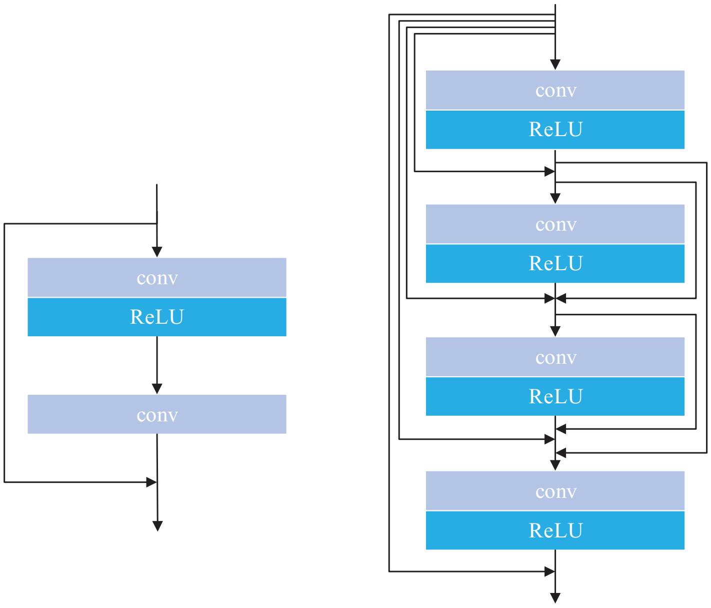
Figure 3: The process of reconstructing secret images through the reconstructed network

Table 2 provides specific parameter details of the network structure.

Additionally, our network architecture integrates Residual Blocks (RB) and improved Dense Blocks (DB), as depicted in Fig. 4. The cascade of feature maps through RBs and DBs aids in the deeper exploration of local features within the stego image and facilitates a form of memory effect for these shallow features. This design enhances the network's feature robustness when confronted with attacks, allowing for the more efficient utilization of these features. Ultimately, the final convolutional layer of the network employs a Sigmoid activation function, effectively mapping the secret image from the feature space back to the conventional image space.

Table 2: The network structure of the reconstructed neural network

Network layer	Channels	Convolutional kernel size	Size of the feature map	Activation function
input	3	–	3 ∗ 128 ∗ 128	–
Conv	3	3 ∗ 3	64 ∗ 64 ∗ 64	ReLU
Conv	64	3 ∗ 3	128 ∗ 64 ∗ 64	ReLU
Conv	128	3 ∗ 3	128 ∗ 64 ∗ 64	ReLU
RB	128	3 ∗ 3	128 ∗ 64 ∗ 64	ReLU
DB	128	3 ∗ 3	128 ∗ 64 ∗ 64	ReLU
RB	128	3 ∗ 3	128 ∗ 64 ∗ 64	ReLU
RB	128	3 ∗ 3	128 ∗ 64 ∗ 64	ReLU
RB	128	3 ∗ 3	128 ∗ 64 ∗ 64	ReLU
Dconv	256	3 ∗ 3	64 ∗ 128 ∗ 128	ReLU
Conv	64	3 ∗ 3	3 ∗ 128 ∗ 128	Sigmoid
output	–	–	3 ∗ 128 ∗ 128	–



(a) Residual Block

(b) Dense Block

Figure 4: The structure diagram of RB and DB

Incorporating these design elements, our reconstruction network is capable of effectively recovering the secret image information from stego images that have been subjected to attacks.

2.3.2 Training the Reconstruction Network

We specifically curated a training set composed of stego images that had been processed with various attack methods. Initially, we synthesized a batch of stego images by combining a set of secret images with a set of content images, all resized to a uniform dimension of 128 × 128 pixels. Subsequently, we conducted a series of attacks on these stego images, including Gaussian noise, salt and pepper noise, JPEG compression, and so on, thereby generating a set of degraded stego images. Based on this, approximately two-thirds of the degraded images were used to form the training set for training the reconstruction network, while the remaining one-third was reserved for testing the network's performance.

Each training session of the reconstruction network focuses on learning how to recover a specific secret image from the stego images. Each stego image input into the reconstruction network contains the style features of the secret image, and the training goal of the network is to accurately recover the secret image from these images. As shown in Eq. (6), we define the loss function of the reconstructed network based on the pixel by pixel Euclidean distance between the final output of the network and the original secret image.

$$MSE = \frac{1}{WH}\sum_{i=1}^{W}\sum_{j=1}^{H}\left(a\left(i,j\right) - a'\left(i,j\right)\right)^2 \tag{6}$$

In this study, the variables W and H represent the horizontal width and vertical height of the secret image, respectively. $a\,(i, j)$ denotes the pixel value of the secret image I_E at the coordinates (i, j), while $a'(i, j)$ signifies the pixel value of the reconstructed secret image I_E at the same coordinates.

We have constructed a pre-trained reconstruction network that is capable of handling stego images that have been subjected to various typical attacks. This training dataset is composed of these compromised stego images, ensuring that the network can effectively recover high-visual-quality hidden images and demonstrate strong robustness against common attack methods.

3 Experimental Results and Analysis

The experimental platform adopts PyTorch 1.9.1, CUDA 11.1 version, programming language Python 3.7, and RTX 3080 graphics card with 12 GB of computing capacity. The content images were selected from the ImageNet [22] dataset, while the secret images were sourced from the Google Open Images [23] dataset.

We meticulously chose 100 content images and 1 secret image from each dataset. Utilizing a Neural Style Transfer algorithm, we generated stego images and subjected them to attacks such as Gaussian noise, ultimately creating 10,000 compromised stego images. These images constituted the dataset for training the reconstruction network. All images were uniformly resized to a dimension of 128 × 128 × 3. In our experiments, we found that after 1000 iterations, the stego images achieved the best match in content with the content images and in style with the secret images. This process ensures that the stego images are visually indistinguishable from regular images.

3.1 Security Analysis

The primary objective of steganography is to achieve the secure and covert transmission of secret information, making security a critical metric for evaluating steganographic algorithms. When transmitting stego images over public channels, we ensure that the content of the secret image cannot be identified by third-party attackers, thereby safeguarding the information's concealment.

Unlike conventional pixel-based steganographic techniques, the method presented in this study does not alter the pixel values of the carrier image. Our approach constructs new stego images by combining the stylistic features of the secret image with the content features of the content image, without revealing the specific content of the secret image. Theoretically, this method can effectively resist attacks from all known steganographic analysis techniques.

To validate the security of our method, we specifically designed experiments that involve a simple linear pixel value differential operation between content images without secret information and stego images containing stylistic information of the secret image. The resulting residual images, as shown in Fig. 5, demonstrate the effectiveness of our approach.

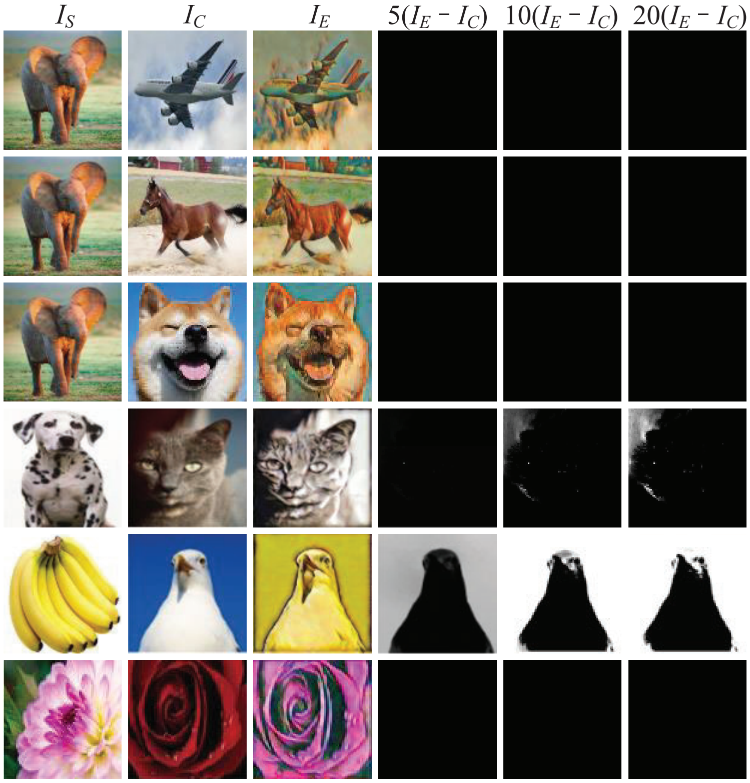


Figure 5: Residual image between stego images and public images

The figure shows the results of enlarging the secret image, content image, generated stego image, and their residual images by 5×, 10×, and 20×. The experimental results show that the residual image does not leak any information about the secret image, thus confirming the security of our proposed steganography model.

3.2 Robustness Analysis

Ensuring the robustness of the steganographic model is crucial [24], as it relates to whether the model can stably reconstruct the secret image in the face of multiple attacks. To test the robustness of our model, we designed a series of experiments, which included applying nine different degrees of attacks on stego images, such as mean filtering, Gaussian filtering, Gaussian noise, median filtering, JPEG compression, salt-and-pepper noise, quantization attacks, and combinations of these attacks.

Fig. 6 presents the effects of these stego images after being attacked, showing in order the images that were not attacked and the images that were subjected to the various attacks mentioned above.

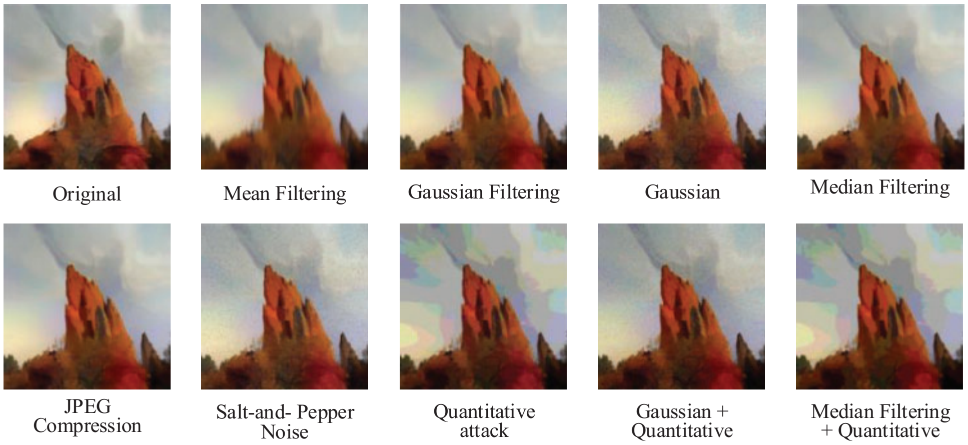


Figure 6: The results of stego images after different attacks

We further selected a total of 18 stego images subjected to various attacks, categorized into two types: roses and mountains, as different secret images. These images were input into the reconstruction network that was pre-trained specifically for the particular secret images. By calculating the PSNR and SSIM values of the reconstructed secret images, we were able to assess the model's robustness, with the results recorded in Table 3. PSNR and SSIM can be used as auxiliary means to evaluate the impact of steganography algorithms on image quality, as well as the degree of modification made by steganalysis methods on images. By monitoring changes in PSNR and SSIM, analysts can have a rough understanding of whether images have been modified and the degree of modification.

Table 3: PSNR (Peak Signal-to-Noise Ratio) and SSIM (Structural Similarity) values of reconstructed secret images

Attack type	Secret image 1		Secret image 2	
	PSNR	SSIM	PSNR	SSIM
Original	39.95	0.998	39.97	0.999
JPEG compression	39.81	0.979	39.85	0.980
Mean filtering	39.18	0.915	39.20	0.910
Quantitative attack	39.25	0.927	39.40	0.936
Gaussian	39.86	0.988	39.81	0.986
Gaussian + Quantitative	38.52	0.868	38.66	0.878
Gaussian filtering	38.86	0.925	39.11	0.902
Median filtering	39.15	0.899	39.01	0.899
Median filtering + Quantitative	38.66	0.854	38.79	0.869

The experimental results show that although quantization attacks may have some impact on the color information of the stego images, resulting in slightly lower PSNR and SSIM values compared to other types of attacks, the reconstruction network, after training with Gaussian noise, quantization attacks, and JPEG compression, can still effectively reconstruct the secret images. For attack types not encountered during the training phase, despite a decline in values, the quality of the reconstructed images remains within an acceptable range.

We believe that this robustness is achieved due to the initialization with the secret image during the generation of the stego images, ensuring that they retain rich secret information even when subjected to attacks. Additionally, training the reconstruction network with stego images that have been attacked allows the network to learn the color style of the stego images with minimal interference, thereby accurately reconstructing the secret images. These factors collectively demonstrate the high robustness of our algorithm in the face of attacks.

3.3 Feasibility Analysis

The successful implementation of a steganographic algorithm hinges on its ability to extract visually satisfactory hidden information from unattacked stego images. As shown in Fig. 7, the first four lines display the original secret image, content image, stego image generated using style transfer technology, and secret image reconstructed by the reconstruction network from top to bottom.

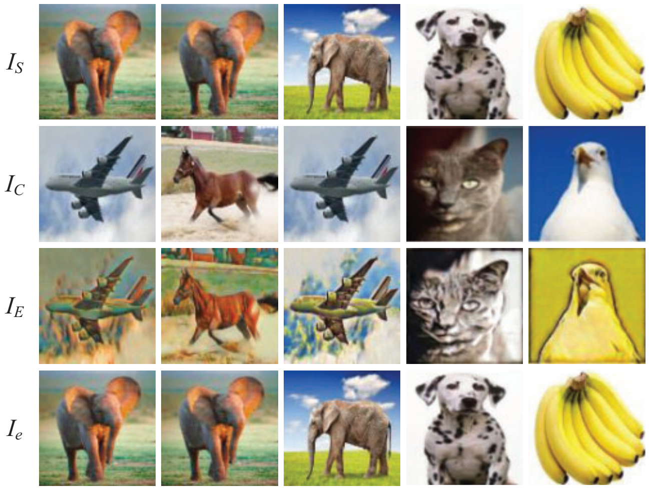


Figure 7: The image effect of the algorithm

By observing Fig. 7, several conclusions can be drawn:

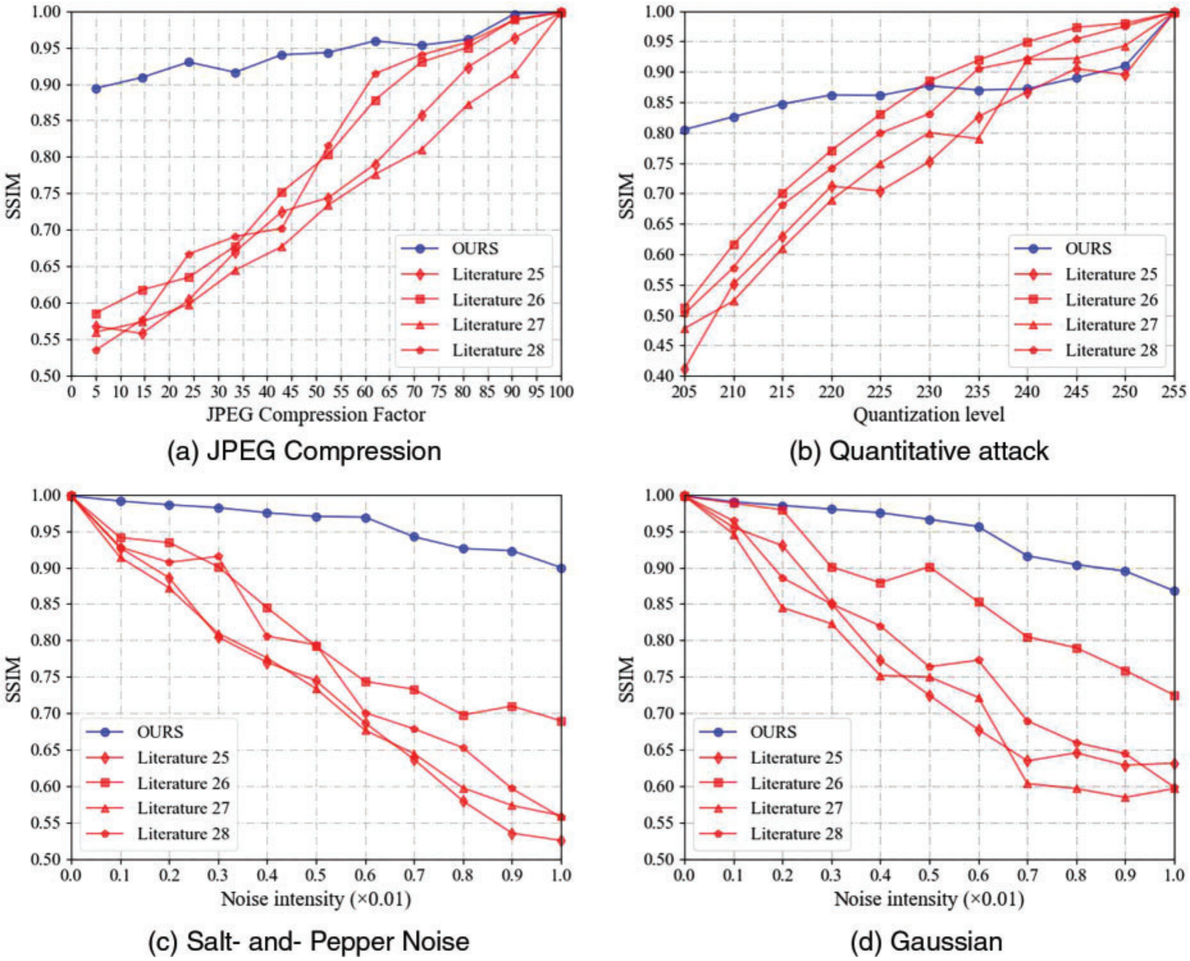
The synthesized stego images contain only the style of the secret image, not the content information. This means that when transmitting stego images over public channels, the content of the secret image cannot be discerned from these images alone, which aligns with the fundamental requirements of steganographic techniques.

The style of the stego images depends on the secret image; if the secret image is the same, the style of the synthesized stego images will also be similar. The content of the stego images, however, is determined by the content image.

To achieve high-quality reconstructed secret images, it is essential to use the target secret image to construct the training dataset for the reconstruction network and define the loss function. Failure to do so will affect the quality of the reconstructed secret images.

3.4 Comparative Analysis

In this study, to thoroughly assess the robustness of the proposed steganographic method, we compared it with four other existing steganographic techniques, which originate from [25–28]. It is noteworthy that the methods from References [25,26], and Reference [28] utilized deep learning-based technologies, while the approach from [27] employed conventional techniques. For the comparative analysis, we used the original secret and content images displayed in Fig. 7 to create stego images and calculated the average SSIM values obtained from reconstructing the secret images through different methods. These results are summarized in Fig. 8.



(a) JPEG Compression

(b) Quantitative attack

(c) Salt- and- Pepper Noise

(d) Gaussian

Figure 8: SSIM value comparison diagram

Through careful observation of the line chart in Fig. 8, we find that in all cases except for quantization attacks, the SSIM values obtained by the method proposed in this paper are significantly better than those of the other four methods. This result highlights the superior performance of the method in this study in terms of resisting attacks and maintaining the stability of steganographic content.

Compared with the deep learning methods of literature [25] and literature [28], the method in this paper uses the color style information of the secret image without modifying the pixel values of the content image when constructing the stego image. The methods of literature [25] and literature [28], on the other hand, directly embed secret information in the carrier image through a neural network. Moreover, while literature [25] and literature [28] directly use the stego image for the extraction network, the method in this paper specifically considers the potential attacks that the stego image may suffer during transmission, using attacked stego images to train the reconstruction network, thereby enhancing the robustness of the method.

Although the method of literature [26] work uses a combination of generator, extractor, and discriminator to generate and extract secret information in the stego image when facing attacks, our reconstruction network, which is trained based on attacked stego images, also shows an advantage in robustness. These comparative results collectively indicate that the steganographic algorithm proposed in this paper has significant advantages in resisting attacks and effectively protecting and reconstructing secret information.

4 Conclusion

In this study, we propose a constructive image steganography technique that relies on style transfer methods to transmit secret images and uses reconstruction networks to reconstruct the secret images. We convert secret information into secret images by mapping dictionaries, and then use style transfer techniques to fuse the style attributes of the secret images with the content attributes of the content images, constructing stego images for information hiding. After a series of extensive experiments, our algorithm has demonstrated its security and robustness in resisting various attack methods. However, when facing quantitative attacks, we found that the robustness of the algorithm is insufficient compared to other types of attacks. Based on this phenomenon, our future research will focus on finding more powerful image features to enhance the algorithm's ability to resist quantitative attacks.

Acknowledgement: The authors are highly thankful to the National Natural Science Foundation of China.

Funding Statement: The authors are highly thankful to the National Natural Science Foundation of China (Nos. 62272478, 61872384, 62172436, 62102451), Natural Science Foundation of Shanxi Province (No. 2023-JC-YB-584), and Engineering University of PAP's Funding for Key Researcher (No. KYGG202011).

Author Contributions: The authors confirm contribution to the paper as follows: study conception and design: Xiong Zhang, Minqing Zhang and Xu'an Wang; data collection: Siyuan Huang; analysis and interpretation of results: Xiong Zhang, Siyuan Huang and Fuqiang Di; draft manuscript preparation: Xiong Zhang. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: Not applicable.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] C. Shen, H. Zhang, D. Feng, Z. Cao, and J. Huang, “Survey of information security,” *Sci. China*, vol. 50, no. 3, pp. 273–298, Jun. 2007. doi: 10.1007/s11432-007-0037-2.

[2] J. Chen, Z. Fu, W. Zhang, X. Cheng, and X. Sun, “Review of image steganalysis based on deep learning,” *J. Softw.*, vol. 32, no. 2, pp. 551–578, Jun. 2021. doi: 10.13328/j.cnki.jos.006135.

[3] J. Wu, Q. Yin, Z. Sheng, W. Lu, J. Huang and B. Li, “Audio multi-view spoofing detection framework based on audio-text-emotion correlations,” *IEEE Trans. Inf. Forensics Secur.*, vol. 19, pp. 7133–7146, Jul. 2024. doi: 10.1109/TIFS.2024.3431888.

[4] J. Li, M. Zhang, K. Niu, Y. Zhang, Y. Ke and X. Yang, “High-security HEVC video steganography method using the motion vector prediction index and motion vector difference,” *Tsinghua Sci. Technol.*, pp. 1–17, Jan. 2024. doi: 10.26599/TST.2024.9010016.

[5] C. Yang, C. Weng, S. Wang, and H. Sun, “Adaptive data hiding in edge areas of images with spatial LSB domain systems,” *IEEE Trans. Inf. Forensics Secur.*, vol. 3, no. 3, pp. 488–497, Aug. 2008. doi: 10.1109/TIFS.2008.926097.

[6] X. Zhang, M. Zhang, X. Wang, W. Jiang, C. Jiang and P. Yang, “Robust information hiding based on neural style transfer with artificial intelligence,” *Comput. Mater. Contin.*, vol. 79, no. 2, pp. 1925–1938, Apr. 2024. doi: 10.32604/cmc.2024.050899.

[7] A. Mohammed, H. Hussein, R. Mstafa, and A. Abdulazeez, “A blind and robust color image watermarking scheme based on DCT and DWT domains,” *Multimed. Tools Appl.*, vol. 82, no. 21, pp. 32855–32881, Mar. 2023. doi: 10.1007/s11042-023-14797-0.

[8] C. Tian, R. Wen, W. Zou, and L. Gong, “Robust and blind watermarking algorithm based on DCT and SVD in the contourlet domain,” *Multimed. Tools Appl.*, vol. 79, no. 11–12, pp. 7515–7541, Mar. 2020. doi: 10.1007/s11042-019-08530-z.

[9] M. Zhou *et al.*, “Deep fourier up-sampling,” 2022, *arXiv:2210.05171*.

[10] Z. Zhou, H. Sun, R. Harit, X. Chen, and X. Sun, “Coverless image steganography without embedding,” in *Int. Conf. Cloud Comput. Secur.*, Nanjing, China, May 2015, vol. 9483, pp. 123–132. doi: 10.1007/978-3-319-27051-7_11.

[11] Z. Zhou, Y. Cao, and X. Sun, “Coverless information hiding based on bag-of-words model of image,” (in Chinese), *J. Appl. Sci.*, vol. 34, no. 5, pp. 527–536, Dec. 2016. doi: 10.3969/j.issn.0255-8297.2016.05.005.

[12] C. Yuan, Z. Xia, and X. Sun, “Coverless image steganography based on SIFT and BOF,” *J. Internet Technol.*, vol. 18, no. 2, pp. 435–442, Mar. 2017. doi: 10.6138/JIT.2017.18.2.20160624c.

[13] Z. Zhou, Y. Cao, M. Wang, E. Fan, and Q. Wu, “Faster-RCNN based robust coverless information hiding system in cloud environment,” *IEEE Access*, vol. 7, pp. 179891–179897, Nov. 2019. doi: 10.1109/ACCESS.2019.2955990.

[14] R. Meng, Z. Zhou, Q. Cui, and X. Sun, “A novel steganography scheme combining coverless information hiding and steganography,” *J. Inf. Hiding Privacy Protection*, vol. 1, no. 1, pp. 43–48, Jan. 2019. doi: 10.32604/jihpp.2019.05797.

[15] Y. Cao, Z. Zhou, C. Yang, and X. Sun, “Dynamic content selection framework applied to coverless information hiding,” *J. Internet Technol.*, vol. 19, no. 4, pp. 1179–1186, Jul. 2018.

[16] C. Wang, Y. Liu, Y. Tong, and J. Wang, “GAN-GLS: Generative lyric steganography based on generative adversarial networks,” *Comput. Mater. Contin.*, vol. 69, no. 1, pp. 1375–1390, Jun. 2021. doi: 10.32604/cmc.2021.017950.

[17] R. Shi, Z. Wang, Y. Hao, and X. Zhang, “Steganography in style transfer,” *IEEE Trans. Artif. Intell.*, pp. 1–12, Mar. 2024. doi: 10.1109/TAI.2024.3379946.

[18] Q. Wang, S. Li, X. Zhang, and G. Feng, “Rethinking neural style transfer: Generating personalized and watermarked stylized images,” in *The 31st ACM Int. Conf. Multimed.*, Ottawa, ON, Canada, Oct. 2023, pp. 6928–6937. doi: 10.1145/3581783.36122.

[19] A. Champandard, “Semantic style transfer and turning two-bit doodles into fine artworks,” *Comput. Res. Repos.*, Mar. 2016. doi: 10.48550/arXiv.1603.01768.

[20] L. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. Yuille, “Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018. doi: 10.1109/TPAMI.2017.2699184.

[21] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” Apr. 2015, *arXiv:1409.1556*.

[22] J. Deng, W. Dong, R. Socher, L. Li, K. Li and F. Li, “ImageNet: A large-scale hierarchical image database,” in *2009 IEEE Conf. Comput. Vis. Pattern Recognit.*, Miami, FL, USA, 2009, pp. 248–255. doi: 10.1109/CVPR.2009.5206848.

[23] A. Kuznetsova *et al.*, “The open images dataset V4: Unified image classification, object detection, and visual relationship detection at scale,” *Int. J. Comput. Vis.*, vol. 128, no. 7, pp. 1956–1981, 2020. doi: 10.1007/s11263-020-01316-z.

[24] Y. Zhang, X. Luo, Y. Guo, C. Qin, and F. Liu, “Multiple robustness enhancements for image adaptive steganography in lossy channels,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 8, pp. 2750–2764, 2020. doi: 10.1109/TCSVT.2019.2923980.

[25] X. Duan, K. Jia, B. Li, D. Guo, E. Zhang and C. Qin, “Reversible image steganography scheme based on a U-Net structure,” *IEEE Access*, vol. 7, pp. 2169–3536, Jan. 2019. doi: 10.1109/ACCESS.2019.2891247.

[26] Q. Li, X. Wang, X. Wang, B. Ma, C. Wang and Y. Shi, “An encrypted coverless information hiding method based on generative models,” *Inf. Sci.*, vol. 553, no. 3, pp. 19–30, Dec. 2020. doi: 10.1016/j.ins.2020.12.002.

[27] I. Kadhim, P. Premaratn, and P. Vial, “Improved image steganography based on super-pixel and coefficient-plane-selection,” *Signal Process.*, vol. 171, Jan. 2020, Art. no. 107481. doi: 10.1016/j.sigpro.2020.107481.

[28] N. Subramanian, I. Cheheb, O. Elharrouss, S. Al-Maadeed, and A. Bouridane, “End-to-end image steganography using deep convolutional autoencoders,” *IEEE Access*, vol. 9, pp. 135585–135593, Sep. 2021. doi: 10.1109/ACCESS.2021.3113953.